On the Least Number of Palindromes Contained in an Infinite Word

Gabriele Fici*

Dipartimento di Matematica e Informatica, Università di Palermo, Italy

Luca Q. Zamboni

Université Claude Bernard Lyon 1, France, and University of Turku, Finland

Abstract

We investigate the least number of palindromic factors in an infinite word. We first consider general alphabets, and give answers to this problem for periodic and non-periodic words, closed or not under reversal of factors. We then investigate the same problem when the alphabet has size two.

Keywords: Combinatorics on words; palindromes.

1. Introduction

In recent years, there has been an increasing interest in the importance of palindromes in mathematics, theoretical computer science and theoretical physics. In particular, one is interested in infinite words containing arbitrarily long palindromes. This stems from their role in the modeling of *quasicrystals* in theoretical physics (see for instance [14, 21]) and also diophantine approximation in number theory (e.g., see [19, 1, 2, 3, 4, 13, 24, 25]).

In [18], X. Droubay, J. Justin and G. Pirillo observed that any finite word w of *length* $|w|$ contains at most $|w|+1$ distinct palindromes (including the empty word). Such words are 'rich' in palindromes, in the sense that they contain the maximum number of different palindromic factors. Accordingly, we say that a finite word w is *rich* if it contains exactly $|w|+1$ distinct palindromes, and we call an infinite word *rich* if all of its factors are rich. In an independent work, P. Ambrož, C. Frougny, Z. Masáková and E. Pelantová [7] have considered the same class of words, which they call *full words* (following earlier work of S. Brlek, S. Hamel, M. Nivat, and C. Reutenauer [10]). Since [18], there is an extensive number of papers devoted to the study of rich words and their generalizations (see [6, 7, 8, 10, 11, 12, 16, 20]).

In this note we consider the opposite question: What is the least number of palindromes which can occur in an infinite word subject to certain constraints? For an infinite word ω, the set $\mathrm{PAL}(\omega)$ of palindromic factors of ω can be finite or infinite (cf. [10]). For instance, in case ω is a Sturmian word, then $\mathrm{PAL}(\omega)$ contains two elements of each odd length and one element of each even length. In fact, this property characterizes Sturmian words (see [15] and [17]). In contrast,

*Corresponding author.
Email addresses: gabriele.fici@unipa.it (Gabriele Fici), lupastis@gmail.com (Luca Q. Zamboni)

Accepted for publication in Theoretical Computer Science *February 7, 2013*

the *paperfolding word* P is an example of an aperiodic uniformly recurrent word containing a finite number of palindromes:

$$P = aabaabbaaabbabbaaabaabbbaabbabbaaaba\cdots$$

It is obtained as the limit of the sequence $(P_n)_{n\geq 0}$ defined recursively by $P_0 = a$ and $P_{n+1} = P_n a \hat{P}_n$ (for $n \geq 0$), where $\hat{P}_n$ is the word obtained from $\tilde{P}_n$ by exchanging a's and b's [5]. J.-P. Allouche showed that the paperfolding word contains exactly 29 palindromes, the longest of which has length 13.

It is easy to see that any uniformly recurrent word ω which contains an infinite number of palindromes must be *closed under reversal*, that is, for every factor $u = u_1u_2\cdots u_n$ of ω, its *reversal* $\tilde{u} = u_n\cdots u_2u_1$ is also a factor of ω. The converse is not true: In fact, J. Berstel, L. Boasson, O. Carton and I. Fagnot [9] exhibited various examples of uniformly recurrent words closed under reversal and containing a finite number of palindromes. The paperfolding word is not closed under reversal, since for example it contains the factor $aaaba$ but not $abaaa$.

If X is a set consisting of infinite words, we set

$$\mathit{MinPal}(X) = \inf\{\#\mathrm{PAL}(\omega) \mid \omega \in X\}.$$

We first show that without restrictions on the cardinality of the alphabet, one has that $\mathit{MinPal} = 4$. That is, for

$$W = \{\omega \in A^{\mathbb{N}} \mid 0 < \#A < \infty\},$$

we have $\mathit{MinPal}(W) = 4$. If in addition one requires that the word be aperiodic, that is, for

$$W_{\mathrm{ap}} = \{\omega \in A^{\mathbb{N}} \mid 0 < \#A < \infty \text{ and } \omega \text{ is aperiodic}\},$$

then $\mathit{MinPal}(W_{\mathrm{ap}}) = 5$. If moreover one requires that the word must be closed under reversal, that is, for

$$W_{\mathrm{cl}} = \{\omega \in A^{\mathbb{N}} \mid 0 < \#A < \infty \text{ and } \omega \text{ is closed under reversal}\},$$

then one still has $\mathit{MinPal}(W_{\mathrm{cl}}) = 5$.

In the case of binary words we show the following:

Theorem 1. *Let A be a set with $\#A = 2$. Then:*

1. $\mathit{MinPal}(A^{\mathbb{N}}) = 9$*, where* $A^{\mathbb{N}}$ *denotes the set of all infinite words on* A*.*
2. $\mathit{MinPal}(A^{\mathbb{N}}_{ap}) = 11$*, where* $A^{\mathbb{N}}_{ap}$ *denotes the set of all aperiodic words in* $A^{\mathbb{N}}$*.*
3. $\mathit{MinPal}(A^{\mathbb{N}}_{cl}) = 13$*, where* $A^{\mathbb{N}}_{cl}$ *denotes the set of all words in* $A^{\mathbb{N}}$ *closed under reversal.*
4. $\mathit{MinPal}(A^{\mathbb{N}}_{ap/cl}) = 13$*, where* $A^{\mathbb{N}}_{ap/cl}$ *denotes the set of all aperiodic words in* $A^{\mathbb{N}}$ *closed under reversal.*

2. Definitions and Notations

Given a finite non-empty set A (called the *alphabet*), we denote by A^* and $A^{\mathbb{N}}$ respectively the set of finite words and the set of (right) infinite words over the alphabet A. Given a finite word $u = a_1a_2\cdots a_n$ with $n \geq 1$ and $a_i \in A$, we denote the length n of u by $|u|$. The *empty word* will be denoted by ε and we set $|\varepsilon| = 0$. We put $A^+ = A^* - \{\varepsilon\}$. For $u, v \in A^+$ we denote by $|u|_v$ the

number of occurrences of v in u. For instance $|0110010|_{01} = 2$. We denote the reverse of u by $\tilde{u}$, i.e., $\tilde{u} = a_n \cdots a_2 a_1$.

Given a finite or infinite word $\omega = \omega_0\omega_1\omega_2\cdots$ with $\omega_i \in A$, we say a word $u \in A^+$ is a *factor* of ω if $u = \omega_i\omega_{i+1}\cdots\omega_{i+n}$ for some natural numbers i and n. We denote by $\mathrm{Fac}(\omega)$ the set of all factors of ω, and by $\mathrm{Alph}(\omega)$ the set of all factors of ω of length 1. Given (non-empty) factors u and v of ω, we say u is a *first return* to v *in* ω if uv is a factor of ω which begins and ends in v and $|uv|_v = 2$. If u is a first return to v in ω then uv is called a *complete first return* to v in ω.

A factor u of ω is called *right special* if both ua and ub are factors of ω for some pair of distinct letters $a, b \in A$. Similarly, u is called *left special* if both au and bu are factors of ω for some pair of distinct letters $a, b \in A$. The factor u is called *bispecial* if it is both right special and left special. For each factor u of ω, we set

$$\omega\big|_u = \{n \in \mathbb{N} \mid \omega_n\omega_{n+1}\cdots\omega_{n+|u|-1} = u\}.$$

We say ω is *recurrent* if for every $u \in \mathrm{Fac}(\omega)$ the set $\omega\big|_u$ is infinite. We say ω is *uniformly recurrent* if for every $u \in \mathrm{Fac}(\omega)$ the set $\omega\big|_u$ is syndedic, i.e., of bounded gap. A word $\omega \in A^{\mathbb{N}}$ is *(purely) periodic* if there exists a positive integer p such that $\omega_{i+p} = \omega_i$ for all indices i, and it is *ultimately periodic* if $\omega_{i+p} = \omega_i$ for all sufficiently large i. For a finite word $u = a_1a_2\cdots a_n$, we call p a *period* of u if $a_{i+p} = a_i$ for every $1 \le i \le n-p$, and we denote by $\pi(u)$ the least period of u. Finally, a word $\omega \in A^{\mathbb{N}}$ is called *aperiodic* if it is not ultimately periodic. Two finite or infinite words are said to be *isomorphic* if the two words are equal up to a renaming of the letters. We denote by $[\omega]$ the set of words that are isomorphic to ω or to $\tilde{\omega}$. Note that any word in $[\omega]$ has the same periods as ω.

We denote by $\mathrm{PAL}(\omega)$ the set of all palindromic factors of ω, i.e., the set of all factors u of ω with $\tilde{u} = u$. We have that $\mathrm{PAL}(\omega)$ contains at least ε and $\mathrm{Alph}(w)$.

3. General alphabets

The following lemma follows from a direct inspection.

Lemma 3.1. *Every word w of length 9 such that $Alph(w) = 2$ contains at least 9 palindromes.*

An application of the previous lemma is the following.

Proposition 3.2. *Every infinite word contains at least 4 palindromes.*

Proof. The empty word and the letters are palindromes. Therefore, if an infinite word ω contains only 3 palindromes then $\mathrm{Alph}(\omega) = 2$. This is in contradiction with Lemma 3.1. □

We have the following characterization of words containing only 4 palindromes.

Proposition 3.3. *If an infinite word contains exactly 4 palindromes, then it is of the form u^∞ where u is of the form $u = abc$ with a, b, and c distinct letters.*

Proof. Let ω be an infinite word containing exactly 4 palindromes. By Lemma 3.1, it follows that $\mathrm{Alph}(\omega) = 3$. Should ω contain a factor of the form aa or of the form aba, then ω would contain at least 5 palindromes. The statement now follows. □

Corollary 3.4. *Every non-periodic infinite word contains at least 5 palindromes.*

In fact, there exist non-periodic uniformly recurrent words containing only 5 palindromes. Let F be the Fibonacci word, that is the word

$$F = abaababaabaababaababaabaab\cdots$$

obtained as the limit of the sequence $(f_n)_{n\geq 0}$, where $f_0 = b$, $f_1 = a$ and $f_{n+1} = f_n f_{n-1}$. The image of the Fibonacci word F under the morphism $\phi : a \mapsto a, b \mapsto bc$,

$$\phi(F) = abcaabcabcaabcaabcabcaabcabca\cdots$$

contains only 5 palindromes, namely: ε, a, b, c and aa. Note that the word $\phi(F)$ is not closed under reversal, since for example it does not contain the reversal of the factor bc.

Berstel et al. [9] exhibited a uniformly recurrent word over a four-letter alphabet closed under reversal and containing only 5 palindromes (the letters and the empty word):

$$\omega = abcdbacdabdcbacdabcdbadcabdcba\cdots$$

defined as the limit of the sequence $(U_n)_{n\geq 0}$, where $U_0 = ab$ and $U_{n+1} = U_n cd\tilde{U}_n$.

4. Binary alphabet

In this section we fix a binary alphabet $A = \{a, b\}$. As a consequence of Lemma 3.1, every infinite word over A contains at least 9 palindromes.

By direct computation, if w is a word over A of length 12, then $\#\mathrm{PAL}(w) \geq 9$ and $\#\mathrm{PAL}(w) = 9$ if and only if $w = u^2$ where $u \in [v]$ and $v = aababb$. Indeed, for each $u \in [v]$ one has

$$\mathrm{PAL}(u^2) = \{\varepsilon, a, b, aa, bb, aba, bab, abba, baab\}.$$

Since no palindrome of length 5 or 6 occurs in u^2, the word u^∞ contains only 9 palindromes. Moreover, for each $u \in [v]$ and $\alpha \in A$, if $\pi(u^2\alpha) \neq 6$, then $u^2\alpha$ contains at least 10 palindromes. So we have:

Proposition 4.1. *Let $v = aababb$. An infinite word over A contains exactly 9 palindromes if and only if it is of the form u^∞ for some $u \in [v]$. In particular it is periodic of period 6.*

We next characterize all binary words containing precisely 10 palindromes. By direct inspection, any word over A of length 14 containing precisely 10 palindromes belongs to one of the following four sets:

1. $T_1 = \{w^2 \mid w \in [av]\}$;
2. $T_2 = \{w^2 \mid w \in [vb]\}$;
3. $T_3 = \{\alpha w^2 \beta \mid \alpha, \beta \in A, w \in [v], \pi(\alpha w^2) \neq 6, \pi(w^2\beta) = 6\}$;
4. $T_4 = \{w^2 \alpha\beta \mid \alpha, \beta \in A, w \in [v], \pi(w^2\alpha) = 6, \pi(w^2\alpha\beta) \neq 6\}$.

Moreover, the length of the longest palindrome in any of the words in the sets T_i is at most 6.

Lemma 4.2. *Let $\gamma \in A$. Then:*

1. *if $w^2 \in T_1$ and $\pi(w^2\gamma) \neq 7$, then $w^2\gamma$ contains 11 palindromes;*
2. *if $w^2 \in T_2$ and $\pi(w^2\gamma) \neq 7$, then $w^2\gamma$ contains 11 palindromes;*

3. *if* $\alpha w^2\beta \in T_3$ *and* $\pi(w^2\beta\gamma) \neq 6$*, then* $w^2\beta\gamma$ *contains* 11 *palindromes;*
4. *if* $w^2\alpha\beta \in T_4$*, then* $w^2\alpha\beta\gamma$ *contains* 11 *palindromes.*

Thus we have:

Proposition 4.3. *An infinite word* w *over* A *contains exactly* 10 *palindromes if and only if* w *is of the form* u^∞ *with* $u \in [av]$ *or* $u \in [vb]$*, or of the form* $\alpha(u)^\infty$ *with* $u \in [v]$ *and* $\alpha \in A$ *such that* αu *does not have period* 6*. In the first case* w *is periodic of period* 7*, while in the second case* w *is ultimately periodic of period* 6*.*

Thus, every aperiodic word over A contains at least 11 palindromes. An example of a uniformly recurrent aperiodic word containing exactly 11 palindromes is the image of the Fibonacci word F under the morphism $\psi : a \mapsto a, b \mapsto abbab$,

$$\psi(F) = aabbabaaabbabaabbabaaabbabaaabbab\cdots$$

The palindromes in $\psi(F)$ are: ε, a, b, aa, bb, aaa, aba, bab, $abba$, $baab$ and $baaab$. Note that the word $\psi(F)$ is not closed under reversal, since, for example, it does not contain the reversal of its factor $abaaa$.

Berstel et al. [9] exhibited an aperiodic uniformly recurrent word closed under reversal and containing precisely 17 palindromes. It is the word obtained from the paperfolding word P by applying the morphism $\tau : a \mapsto ab, b \mapsto ba$:

$$\tau(P) = ababbaababbabaababababbabaabbabaababab\cdots$$

In the next section, we show that the least number of palindromes which can occur in an infinite binary word closed under reversal is 13.

Rather than limiting the total number of palindromic factors, we consider the problem of limiting the length of the longest palindromic factor. In this case we have:

Proposition 4.4. *Every infinite binary word contains a palindromic factor of length greater than* 3*. There exist infinite binary words containing no palindromic factor of length greater than* 4*, but every such word is ultimately periodic. There exists a uniformly recurrent aperiodic binary word (closed under reversal) whose longest palindromic factor has length* 5*.*

Proof. Let $\omega \in \{a,b\}^{\mathbb{N}}$. We will show that ω contains a palindromic factor of length greater than 3. Let T denote the shift map, that is, $T\omega$ is the word whose i-th letter is ω_{i+1}. If $aaa \in \mathrm{Fac}(T\omega)$, then

$$\{aaaa, baaab\} \cap \mathrm{Fac}(\omega) \neq \emptyset.$$

Thus we can assume that neither aaa nor bbb is a factor of $T\omega$. If neither aa nor bb occurs in $T\omega$, then $T\omega = (ab)^\infty$ or $T\omega = (ba)^\infty$. In either case $T\omega$ contains the palindrome $ababa$. Thus, without loss of generality we can assume that aa occurs in $T\omega$. If we now consider all the possible right extensions of aa which avoid aaa and bbb, we find that each terminates in a palindrome of length 4 or 5:

$$\begin{cases} aabaa \\ aababa \\ aababba \\ aabba \end{cases}$$

Next, suppose ω contains no palindromic factor of length greater than 4. We will show that ω is ultimately periodic and give an example of such a word. If $T\omega$ contains $aaaa$, then $baaaa \in \mathrm{Fac}(\omega)$ which implies that

$$\{aaaaa, baaaab\} \cap \mathrm{Fac}(\omega) \neq \emptyset,$$

a contradiction. Thus we can assume that $aaaa \notin \mathrm{Fac}(T\omega)$. If $aaa \in \mathrm{Fac}(T^2\omega)$, then $baaa \in \mathrm{Fac}(T\omega)$, which implies that $baaab \in \mathrm{Fac}(\omega)$, a contradiction. Thus neither aaa nor bbb occurs in $T^2\omega$. If neither aa nor bb occurs in $T^2\omega$, we would have that $T^2\omega = (ab)^\infty$ or $T^2\omega = (ba)^\infty$, a contradiction since each contains $ababa$. Thus, without loss of generality we can assume that aab occurs in $T^2\omega$. It is readily verified that the only possible first returns to aab in ω are

$$\begin{cases} aababb \\ aabbab \end{cases}$$

If $aabbab$ occurs in ω, then ω has a tail of the form $(aabbab)^\infty$. If $aabbab$ does not occur in ω, the ω has a tail of the form $(aababb)^\infty$. In either case ω is ultimately periodic. It is readily verified that $(aabbab)^\infty$ has no palindromic factor of length greater than 4.

Finally, we show the existence of a uniformly recurrent aperiodic binary word ω (closed under reversal) whose longest palindromic factor has length 5. Set $U_0 = aabb$ and for $n \geq 0$,

$$\begin{cases} U_{2n+1} = U_{2n}ab\tilde{U}_{2n}; \\ U_{2n} = U_{2n-1}ba\tilde{U}_{2n-1}. \end{cases}$$

Then U_n is a prefix of U_{n+1} for each $n \geq 0$ and we set

$$\omega = \lim_{n\to\infty} U_n.$$

Then, by construction, ω is closed under reversal and is uniformly recurrent (in fact, the recursive definition of ω shows that each prefix U_n occurs in ω with bounded gap). Now, a straightforward verification shows that

$$\mathrm{PAL}(U_2) = \{\varepsilon, a, b, aa, bb, aaa, aba, bab, bbb, abba, baab, aabaa, abbba, baaab, bbabb\}.$$

We note that $\mathrm{PAL}(\tilde{U}_n) = \mathrm{PAL}(U_n)$ for each $n \geq 0$. We prove by induction on $n \geq 2$ that no other palindrome occurs in U_n. From the above equality, we have that the result holds for $n = 2$. Now, suppose that $n \geq 2$ and $\#\mathrm{PAL}(U_n) = 15$. We will show that $\#\mathrm{PAL}(U_{n+1}) = 15$. For $n \geq 2$, we can write $U_n = U_1 t_n \tilde{U}_1$ and

$$U_{n+1} = \begin{cases} U_1 t_n \tilde{U}_1 ab U_1 \tilde{t}_n \tilde{U}_1 & \text{for } n \text{ even;} \\ U_1 t_n \tilde{U}_1 ba U_1 \tilde{t}_n \tilde{U}_1 & \text{for } n \text{ odd.} \end{cases}$$

Considering that $|U_1| = 10$, if U_{n+1} contained a palindrome v of length 6 or 7, then either v would be contained in U_n or in U_2. Thus, $\#\mathrm{PAL}(U_{n+1}) = 15$. Finally, by Lemma 5.2, we deduce that ω is aperiodic. □

5. The case of binary words closed under reversal

In this section we will prove the following:

Theorem 5.1. *Let $A = \{a, b\}$, and let $A^{\mathbb{N}}_{cl}$ (respectively, $A^{\mathbb{N}}_{ab/cl}$) denote the set of all infinite words in $A^{\mathbb{N}}$ closed under reversal (respectively, the set of all aperiodic words in $A^{\mathbb{N}}$ closed under reversal). Then $MinPal(A^{\mathbb{N}}_{cl}) = MinPal(A^{\mathbb{N}}_{ab/cl}) = 13$.*

Note that since there exist aperiodic binary words closed under reversal containing a finite number of palindromic factors (see for instance [9] or Lemma 5.4), we have that

$$MinPal(A^{\mathbb{N}}_{\mathrm{cl}}) \leq MinPal(A^{\mathbb{N}}_{\mathrm{ab/cl}}) < +\infty.$$

Our proof will involve some intermediate lemmas and a case-by-case analysis. We begin with some general remarks concerning words closed under reversal. The following lemma is probably well known but we include it here for the sake of completeness:

Lemma 5.2. *Suppose $\omega \in A^{\mathbb{N}}$ is closed under reversal. Then ω is recurrent. Hence ω is either aperiodic or (purely) periodic. In the latter case $\#PAL(w) = +\infty$.*

Proof. Let u be a prefix of ω. Then $\tilde{u}$ occurs in ω followed by some letter $\alpha \in A$. Then αu is a factor of ω, which means that u occurs at least twice in ω. This proves that ω is recurrent. If ω is ultimately periodic, meaning $\omega = vu^\infty$ for some u and v, then as ω is recurrent it follows that u is a factor of v^∞, which implies that ω is purely periodic. Finally, it remains to prove that if $\omega = u^\infty$ for some factor u of ω, then ω contains an infinite number of palindromic factors, or equivalently that for each M there exists a palindromic factor v of ω with $|v| \geq M$. So let M be given and pick a positive integer n such that $n|u| \geq M$. Since u^n is a factor of ω it follows that $\tilde{u}^n$ is a factor of ω. This implies that $\tilde{u}^n$ occurs in u^{n+1}. Thus there exists a factor v of ω with $n|u| \leq |v| < (n+1)|u|$ which begins in u^n and ends in $\tilde{u}^n$. Hence v is a palindrome. □

Corollary 5.3. *$MinPal(A^{\mathbb{N}}_{cl}) = MinPal(A^{\mathbb{N}}_{ab/cl})$.*

Proof. We already noticed that $MinPal(A^{\mathbb{N}}_{\mathrm{cl}}) \leq MinPal(A^{\mathbb{N}}_{\mathrm{ab/cl}}) < +\infty$. Let $\omega \in A^{\mathbb{N}}_{\mathrm{cl}}$ with $\#\mathrm{PAL}(\omega) = MinPal(A^{\mathbb{N}}_{\mathrm{cl}})$. Since $MinPal(A^{\mathbb{N}}_{\mathrm{cl}}) < +\infty$, it follows from Lemma 5.2 that ω is aperiodic and hence $\omega \in A^{\mathbb{N}}_{\mathrm{ab/cl}}$, whence $MinPal(A^{\mathbb{N}}_{\mathrm{ab/cl}}) \leq \#\mathrm{PAL}(\omega) = MinPal(A^{\mathbb{N}}_{\mathrm{cl}})$. □

We begin by showing that 13 is an upper bound for $MinPal(A^{\mathbb{N}}_{\mathrm{cl}})$.

Lemma 5.4. *$MinPal(A^{\mathbb{N}}_{cl}) \leq 13$.*

Proof. Set $U_0 = abaabbabaaabbaaba$. For $n \geq 0$ define

$$\begin{cases} U_{2n+1} = U_{2n}\, bbaa\, \tilde{U}_{2n}; \\ U_{2n+2} = U_{2n+1}\, aabb\, \tilde{U}_{2n+1}. \end{cases}$$

The first few values of U_n are as follows:

$U_1 = abaabbabaaabbaaba\; bbaa\; abaabbaaababbaaba$

$U_2 = abaabbabaaabbaababbaaabaabbaaababbaaba\; aabb\; abaabbabaaabbaabaaabbabaabbaaababbaaba.$

We note that U_n is a prefix of U_{n+1} for each $n \geq 0$. Let

$$\omega = \lim_{n \to +\infty} U_n.$$

It is clear by construction that ω is closed under reversal. It is readily verified that

$$\mathrm{PAL}(U_2) = \{\varepsilon, a, aa, aaa, aabaa, aabbaa, aba, abba, b, baaab, baab, bab, bb\}.$$

Hence $\#\mathrm{PAL}(U_2) = 13$.

We prove by induction on $n \geq 2$ that $\#\mathrm{PAL}(U_n) = 13$. The above equality shows that this is true for $n = 2$. For every $n \geq 2$, we can write $U_n = U_1 t_n \tilde{U_1}$ and

$$U_{n+1} = \begin{cases} U_1 t_n \tilde{U_1} bbaa U_1 \tilde{t}_n \tilde{U_1} & \text{for } n \text{ even;} \\ U_1 t_n \tilde{U_1} aabb U_1 \tilde{t}_n \tilde{U_1} & \text{for } n \text{ odd.} \end{cases}$$

Since $|U_1| = 38$, a palindrome of length smaller than or equal to 8 which occurs in U_{n+1} must either occur in U_n or in U_2. The result now follows from the induction hypothesis. □

Set

$$\Omega = \{\omega \in \{a, b\} \mid \omega \,\text{is closed under reversal and}\, \#\mathrm{PAL}(\omega) = \mathit{MinPal}(A_{\mathrm{cl}}^{\mathbb{N}})\}.$$

For $\omega \in \Omega$ and $x \in \{a, b\}$ we put

$$N_x(\omega) = \max\{k \mid x^k \,\text{is a factor of}\, \omega\}.$$

Since $\mathit{MinPal}(A_{\mathrm{cl}}^{\mathbb{N}}) < +\infty$ both N_a and N_b are finite.

Proof of Theorem 5.1. Fix $\omega \in \Omega$. We must show that $\#\mathrm{PAL}(\omega) \geq 13$. We will make use of the following lemma:

Lemma 5.5. *Let k be a positive integer. If a^k (respectively b^k) is a factor of ω, then so is ba^kb (respectively ab^ka).*

Proof. Suppose to the contrary that a^k is a factor of ω but not ba^kb. By Lemma 5.2, ω is recurrent and hence $ba^{N_a(\omega)}b$ is a factor of ω. Hence $1 \leq k < N_a(\omega)$. Let ω' be a tail of ω beginning in b. Let $\nu \in \{a, b\}^{\mathbb{N}}$ be the word obtained from ω' by replacing all occurrences of ba^jb in ω' by $ba^{j-1}b$ for each $k + 1 \leq j \leq N_a(\omega)$. Thus $N_a(\nu) = N_a(\omega) - 1$. It is readily verified that ν is closed under reversal. Moreover, to every palindrome v in ν corresponds a unique palindrome w in ω obtained from v by increasing the a runs in v of length $\geq k$ by one unit and leaving the other a runs the same. This defines an injection $\phi : \mathrm{PAL}(\nu) \hookrightarrow \mathrm{PAL}(\omega)$ which is not a surjection since in particular a^k is not in the image of ϕ. Thus $\#\mathrm{PAL}(\nu) < \#\mathrm{PAL}(\omega)$, contradicting that ω had the least number of palindromic factors amongst all binary words closed under reversal. □

Lemma 5.6. *If $\#PAL(\omega) \leq 12$, then a^2 and b^2 are factors of ω.*

Proof. Since $\#\mathrm{PAL}(\omega) \leq 12$, no factor of ω of length 12 is rich. There are 850 binary non-rich words u of length 12. For each such u we compute $\mathrm{PAL}(u)$. By computer verification, we observe that the only cases when $\mathrm{PAL}(u)$ does not contain both a^2 and b^2 is when $\mathrm{PAL}(u)$ is equal to one of the following 4 sets:

$$\begin{aligned}
A &= \{\varepsilon, a, aba, abba, abbba, b, bab, babbab, babbbab, bb, bbabb, bbb\}; \\
B &= \{\varepsilon, a, aba, abba, b, bab, babab, babbab, bb, bbababb, bbabb, bbb\}; \\
C &= \{\varepsilon, b, bab, baab, baaab, a, aba, abaaba, abaaaba, aa, aabaa, aaa\}; \\
D &= \{\varepsilon, b, bab, baab, a, aba, ababa, abaaba, aa, aababaa, aabaa, aaa\}.
\end{aligned}$$

For instance, if $u = aaababaabaaa$ then $\mathrm{PAL}(u) = D$. So either $\mathrm{PAL}(\omega)$ contains both a^2 and b^2 or $\mathrm{PAL}(\omega)$ contains one of A, B, C, or D. Note that each of the above sets is of cardinality 12. By Lemma 5.5, if $B \subseteq \mathrm{PAL}(\omega)$, then $\mathrm{PAL}(\omega)$ also contains $abbba$ and hence $\#\mathrm{PAL}(\omega) \geq 13$, a contradiction. A similar argument shows that $\mathrm{PAL}(\omega)$ cannot contain D. Next, suppose $A \subseteq \mathrm{PAL}(\omega)$ and let us consider the possible right extensions of the palindrome $babbbab$ which avoid a^2. We will put a dot (.) to designate positions of choice. They are: $babbbab.ab$ (which yields a 13th palindrome $babab$), $babbbab.b.ab.ab$ (which yields a 13th palindrome $babab$), $babbbab.b.ab.b$ (which yields a 13th palindrome $bbabbabb$), and finally $babbbab.b.b$ (which yields a 13th palindrome $bbbabbb$). In either case $\#\mathrm{PAL}(\omega) \geq 13$, a contradiction. A similar argument shows that $\mathrm{PAL}(\omega)$ cannot contain C. Thus, if $\#\mathrm{PAL}(\omega) \leq 12$ then both a^2 and b^2 are factors of ω. □

In view of Lemma 5.6, we can suppose that both a^2 and b^2 belong to $\mathrm{PAL}(\omega)$ and hence, by Lemma 5.5,

$$\{\varepsilon, a, b, aa, bb, aba, bab, abba, baab\} \subseteq \mathrm{PAL}(\omega).$$

If ω contains no palindromic factor of length greater than 4, then neither aaa nor bbb is a factor of ω (for otherwise by Lemma 5.5 either $baaab$ or $abbba$ is a factor of ω). Hence we would have

$$\{\varepsilon, a, b, aa, bb, aba, bab, abba, baab\} = \mathrm{PAL}(\omega).$$

This implies that $ababb$ is a factor of ω. But then the only complete first return to $ababb$ is $ababbaababb$, which would imply that ω is periodic, a contradiction. Thus, ω must contain a palindromic factor of length 5 or of length 6.

Case 1: ω contains a palindromic factor of length 5.

Without loss of generality we can suppose

$$\{aabaa, babab, baaab, aaaaa\} \cap \mathrm{PAL}(\omega) \neq \emptyset.$$

Case 1.1: $aabaa \in \mathrm{PAL}(\omega)$. Thus $\#\mathrm{PAL}(\omega) \geq 10$. By considering the possible bilateral extensions of $aabaa$, we have

$$\{aaabaaa, aaabaab, baabaab\} \cap \mathrm{Fac}(\omega) \neq \emptyset.$$

Case 1.1.1: $aaabaaa \in \mathrm{Fac}(\omega)$. This gives rise to 3 additional palindromes: $aaabaaa, aaa, baaab$ (where $baaab$ is a consequence of aaa and Lemma 5.5). Hence $\#\mathrm{PAL}(\omega) \geq 13$.

Case 1.1.2: $aaabaab \in \mathrm{Fac}(\omega)$. This gives rise to 2 additional palindromes: $aaa, baaab$ so that $\#\mathrm{PAL}(\omega) \geq 12$. If $aaaa \in \mathrm{Fac}(\omega)$, then $\#\mathrm{PAL}(\omega) \geq 13$. So we can assume that $aaaa \notin \mathrm{Fac}(\omega)$, in which case $abaaab \in \mathrm{Fac}(\omega)$ since $aaabaab$ occurs in ω preceded by b and ω is closed under reversal. We leave the following technical claim for the reader:

Claim 5.7. *Under the conditions of Case 1.1.2, either $\#PAL(\omega) \geq 13$, or every complete first return to abaaab is of the form $abaaab\,(babaab)^n b\,abaaab$ for $n \geq 0$.*

This implies that $aabb$ is a factor of ω but not $bbaa$, a contradiction.

Case 1.1.3: $baabaab \in \mathrm{Fac}(\omega)$. In this case $\#\mathrm{PAL}(\omega) \geq 11$. We leave the following technical claim for the reader:

Claim 5.8. *Under the conditions of Case 1.1.3, either* $\#PAL(\omega) \geq 13$, *or every complete first return to* $baabaab$ *is either of the form* $baabaab\,(babaab)^n\,aab$ *or of the form* $baabaab\,(abbaab)^n aab$ *for* $n \geq 1$.

Since ω is closed under reversal, both forms must actually occur. But the switch from one form to the other will produce two new palindromes: $abaabaaba$, $babaabaabab$. In either case $\#\mathrm{PAL}(\omega) \geq 13$. This completes Case 1.1.

Case 1.2: $babab \in \mathrm{PAL}(\omega)$. Thus $\#\mathrm{PAL}(\omega) \geq 10$. By considering the possible bilateral extensions of $babab$, we have

$$\{abababa, abababb, bbababb\} \cap \mathrm{Fac}(\omega) \neq \emptyset.$$

Case 1.2.1: $abababa \in \mathrm{Fac}(\omega)$. This gives rise to 2 additional palindromes: $ababa, abababa$ so that $\#\mathrm{PAL}(\omega) \geq 12$. But then every bilateral extension adds a 13th palindrome: either $aabababaa$ or $bababab$.

Case 1.2.2: $abababb \in \mathrm{Fac}(\omega)$. This gives rise to the additional palindrome $ababa$ so that $\#\mathrm{PAL}(\omega) \geq 11$. We leave the following technical claim for the reader:

Claim 5.9. *Under the conditions of Case 1.2.2, either* $\#PAL(\omega) \geq 13$, *or every complete first return to* $ababa$ *is either of the form* $ababa\,(bbaaba)^n\,ba$ *or of the form* $ababa\,(abbaba)^n ba$ *for* $n \geq 0$.

Since ω is closed under reversal, both forms must actually occur. But the switch from one form to the other will produce two new palindromes: $aababaa$, $baababaab$. In either case $\#\mathrm{PAL}(\omega) \geq 13$.

Case 1.2.3: $bbababb \in \mathrm{Fac}(\omega)$. This gives rise to the additional palindrome $bbababb$, so that $\#\mathrm{PAL}(\omega) \geq 11$. If $bbb \in \mathrm{Fac}(\omega)$, then $\mathrm{PAL}(\omega)$ would contain 2 additional palindromes (namely, bbb and $abbba$). So we can assume that $bbb \notin \mathrm{Fac}(\omega)$, in which case $abbababba \in \mathrm{Fac}(\omega)$, so that $\#\mathrm{PAL}(\omega) \geq 12$. If $abbababba$ extends on either side by b, we would get the 13th palindrome $babbab$. Otherwise, $aabbababbaa \in \mathrm{Fac}(\omega)$. In either case, $\#\mathrm{PAL}(\omega) \geq 13$. This completes Case 1.2.

Case 1.3: $baaab \in \mathrm{PAL}(\omega)$. In this case

$$\{\varepsilon, a, b, aa, bb, aba, bab, abba, baab, aaa, baaab\} \subseteq \mathrm{PAL}(\omega),$$

and hence $\#\mathrm{PAL}(\omega) \geq 11$. If either $aaaa$ or bbb is a factor of ω, this would give rise to 2 additional palindromes, whence $\#\mathrm{PAL}(\omega) \geq 13$. So we can assume that $aaaa$ and bbb are not factors of ω and in view of cases 1.1 and 1.2, that $baaab$ is the only palindromic factor of ω of length 5. By considering the possible bilateral extensions of $baaab$, we have

$$\{bbaaabb, abaaaba, abaaabb\} \cap \mathrm{Fac}(\omega) \neq \emptyset.$$

Case 1.3.1: $bbaaabb \in \mathrm{Fac}(\omega)$. But then so is $abbaaabba$ whence $\#\mathrm{PAL}(\omega) \geq 13$.

Case 1.3.2: $abaaaba \in \mathrm{Fac}(\omega)$. So $\#\mathrm{PAL}(\omega) \geq 12$. But any bilateral extension of $abaaaba$ adds a 13th palindrome: either $aabaa$ or $babaaabab$. In either case $\#\mathrm{PAL}(\omega) \geq 13$.

Case 1.3.3: $abaaabb \in \mathrm{Fac}(\omega)$. So $\#\mathrm{PAL}(\omega) \geq 11$. We leave the following technical claim for the reader:

Claim 5.10. *Under the conditions of Case 1.3.3 (which include that aaaa and bbb are not factors of* ω *and that baaab is the only palindromic factor of* ω *of length* 5) *either* $\#PAL(\omega) \geq 13$, *or every first return to baaab is of one of* 4 *types:*

- $x_n = baaab(baabab)^n$ *for some* $n \geq 1$*;*
- $y_n = baaab(babaab)^n$ *for some* $n \geq 1$*;*
- $w_n = baaab(abbaab)^n ab$ *for some* $n \geq 0$*;*
- $z_n = baaab(babaab)^n ba$ *for some* $n \geq 0$.

We now consider two consecutive first returns to *baaab* in ω. If any combination from the following set should occur:

$$\{x_n x_m, x_n y_m, x_n z_m, y_n x_m, y_n y_m, y_n z_m, w_n x_m, w_n y_m, w_n z_m\},$$

then ω would contain 2 additional palindromic factors: *bbaaabb*, *abbaaabba*. So, either $\#\mathrm{PAL}(\omega) \geq 13$ or none of the above combinations occurs in ω. But if none of the above combinations occurs in ω, since ω is closed under reversal, this implies that w_n does not occur in ω (since w_n can only be followed by w_m), and hence neither does z_n. Since x_n and y_n can only be followed by w_m, it follows that x_n and y_n also do not occur in ω. Since ω is recurrent, some first return to *baaab* must occur. Thus, even in this case we have $\#\mathrm{PAL}(\omega) \geq 13$. This completes Case 1.3.

Case 1.4: $aaaaa \in \mathrm{PAL}(\omega)$. In this case

$$\{\varepsilon, a, b, aa, bb, aaa, aba, bab, aaaa, abba, baab, aaaaa, baaab, baaaab, baaaaab\} \subseteq \mathrm{PAL}(\omega),$$

whence $\#\mathrm{PAL}(\omega) \geq 15$. This completes Case 1.

Case 2: ω does not contain any palindromic factors of length 5.

In this case, neither *aaa* nor *bbb* is a factor of ω (for otherwise by Lemma 5.5 ω would contain a palindrome of length 5). But in view of Proposition 4.4, ω must contain a palindromic factor of length 6. Without loss of generality this implies that

$$\{aabbaa, babbab\} \cap \mathrm{PAL}(\omega) \neq \emptyset.$$

Case 2.1: $aabbaa \in \mathrm{PAL}(\omega)$. In this case, $baabbaab \in \mathrm{PAL}(\omega)$ so that $\#\mathrm{PAL}(\omega) \geq 11$. We now consider the possible bilateral extensions of *baabbaab*. The extension *abaabbaaba* gives rise to 2 additional palindromes: *abaabbaaba*, *babaabbaabab* where the second follows from the fact that ω does not contain any palindromic factors of length 5. The extension *abaabbaabb* gives rise to 2 additional palindromes: *bbaabb*, *abbaabba*. Finally, the last extension *bbaabbaabb* gives rise to 2 additional palindromes: *bbaabb*, *bbaabbaabb*. In either case, $\#\mathrm{PAL}(\omega) \geq 13$.

Case 2.2: $babbab \in \mathrm{PAL}(\omega)$. Since ω contains no palindromic factor of length 5, the only bilateral extension of *babbab* is *ababbaba*. So $\#\mathrm{PAL}(\omega) \geq 11$. Again, since ω contains no palindromic factor of length 5, the only bilateral extension of *ababbaba* is *aababbabaa*, which gives rise to 2 additional palindromes: *aababbabaa*, *baababbabaab*. Thus, $\#\mathrm{PAL}(\omega) \geq 13$. This completes Case 2 and the proof of Theorem 5.1. □

Remark 5.11. *One can wonder what happens for infinite words that are generated by morphisms. Actually, Tan [26] proved that if ω is the fixed point of a primitive morphism, then ω is closed under reversal if and only if $\#PAL(\omega) = \infty$.*

6. Acknowledgements

The second author is partially supported by a FiDiPro grant (137991) from the Academy of Finland, and by ANR grant *SUBTILE*.

References

[1] B. Adamczewski, Y. Bugeaud, Palindromic continued fractions, *Ann. Inst. Fourier* **57** (2007) p. 1557–1574.

[2] B. Adamczewski, Y. Bugeaud, Transcendence measure for continued fractions involving repetitive or symmetric patterns, *J. Eur. Math. Soc.* **12** (2010) p. 883–914.

[3] B. Adamczewski, Y. Bugeaud, On the Littlewood conjecture in simultaneous Diophantine approximation, *J. London Math. Soc.* **73** (2006) p. 355–366.

[4] B. Adamczewski, J.-P. Allouche, Reversal and palindromes in continued fractions, *Theoret. Comput. Sci.* **320** (2007) 220–237.

[5] J.-P. Allouche, Schrödinger operators with Rudin-Shapiro potentials are not palindromic, *J. Math. Phys., Special Issue "Quantum Problems in Condensed Matter Physics"* **38** (1997) p. 1843–1848.

[6] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, *Theoret. Comput. Sci.* **292** (2003) p. 9–31.

[7] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, *Ann. Inst. Fourier (Grenoble)* **56** (2006) p. 2131–2160.

[8] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, *Theoret. Comput. Sci.* **380** (2007) p. 266–275.

[9] J. Berstel, L. Boasson, O. Carton and I. Fagnot, Infinite words without palindrome, *CoRR abs/0903.2382*, (2009).

[10] S. Brlek, S. Hamel, M. Nivat and C. Reutenauer. On The Palindromic Complexity Of Infinite Words, *Internat. J. Found. Comput. Sci.* **15** (2004) p. 293–306.

[11] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, *Adv. Appl. Math.* **42** (2009) p. 60–74.

[12] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A new characteristic property of rich words, *Theoret. Comput. Sci.* **410** (2009) p. 2860–2863.

[13] Y. Bugeaud, M. Laurent, Exponents of Diophantine and Sturmian continued fractions, *Ann. Inst. Fourier* **55** (2005) p. 773–804.

[14] D. Damanik, L.Q. Zamboni, Combinatorial properties of Arnoux-Rauzy subshifts and applications to Schrödinger operators, *Rev. Math. Phys.* **15** (2003) p. 745–763.

[15] A. de Luca, Sturmian words: structure, combinatorics and their arithmetics, *Theoret. Comput. Sci.* 183 (1997) 45–82.

[16] A. de Luca, A. Glen, L.Q. Zamboni, Rich, Sturmian, and trapezoidal words, *Theoret. Comput. Sci.* **407** (2008) p. 569–573.

[17] X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.* **223** (1999) p. 73–85.

[18] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.* **255** (2001) p. 539–553.

[19] S. Fischler, Palindromic prefixes and diophantine approximation, *Monatsh. Math.* **151** (2007) p. 11–37.

[20] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, *European J. Combinatorics*, **30** (2009) p. 510–531.

[21] A. Hof, O. Knill, B. Simon, Singular continuous spectrum for palindromic Schrödinger operators, *Commun. Math. Phys.* **174** (1995) p. 149–159.

[22] M. Lothaire, *Combinatorics On Words*, vol. 17 of *Encyclopedia of Mathematics and its Applications*, Addison-Wesley, Reading, Massachusetts, 1983.

[23] M. Lothaire, *Algebraic Combinatorics On Words*, vol. 90 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, U.K., 2002.

[24] D. Roy, Approximation to real numbers by cubic algebraic integers, II, *Ann. of Math.* **158** (2003) p. 1081–1087.

[25] D. Roy, Approximation to real numbers by cubic algebraic integers, I, *Proc. London Math. Soc.* **88** (2004) p. 42–62.

[26] B. Tan. Mirror substitutions and palindromic sequences, *Theoret. Comput. Sci.* **389** (2007) p. 118–124.